



# Gotcha-I: A Multiview Human Videos Dataset

Paola Barra<sup>1</sup>, Carmen Bisogni<sup>1</sup>(✉), Michele Nappi<sup>1</sup>, David Freire-Obregón<sup>2</sup>,  
and Modesto Castrillón-Santana<sup>2</sup>

<sup>1</sup> Università degli Studi di Salerno, 84084 Salerno, Italy

{pbarra,cbisogni,mnappi}@unisa.it

<sup>2</sup> Universidad de Las Palmas de Gran Canaria (ULPGC), 35017 Gran Canaria, Spain

{david.freire,modesto.castrillon}@ulpgc.es

**Abstract.** The growing need of security in large open spaces led to the need to use video capture of people in different context and illumination and with multiple biometric traits as head pose, body gait, eyes, nose, mouth, and further more. All these traits are useful for a multibiometric identification or a person re-identification in a video surveillance context. Body Worn Cameras (BWCs) are used by the police of different countries all around the world and their use is growing significantly. This raises the need to develop new recognition methods that consider multibiometric traits on person re-identification. The purpose of this work is to present a new video dataset called Gotcha-I. This dataset has been obtained using more mobile cameras to adhere to the data of BWCs. The dataset includes videos from 62 subjects in indoor and outdoor environments to address both security and surveillance problem. During these videos, subjects may have a different behavior in videos such as freely, path, upstairs, avoid the camera. The dataset is composed by 493 videos including a set of 180° videos for each face of the subjects in the dataset. Furthermore, there are already processed data, such as: the 3D model of the face of each subject with all the poses of the head in pitch, yaw and roll; and the body keypoint coordinates of the gait for each video frame. It's also shown an application of gender recognition performed on Gotcha-I, confirming the usefulness and innovativeness of the proposed dataset.

**Keywords:** Dataset · Biometric · Face · Gait · Head pose estimation · Mobile device · Body worn cameras

## 1 Introduction

For security purposes, there are various types of videos to examine with different devices, from different perspectives and in different environments. This type of data increasingly comes from mobile devices such as cameras worn by the body by UK police officers. Body worn cameras are evolving the information consumed by the different security agencies. In addition to providing important information on the health of the agents, the main use of these cameras is to

record videos useful in dangerous situations to perform subject and action recognition. To test algorithms suitable for data like these, we created Gotcha-I. Our proposal contains videos with different modes. The subjects in the videos are both cooperative and non-cooperative, in order to simulate the user's attempts to avoid the camera. They move along a path or freely and have been captured in different lighting environments. One of the innovations made by Gotcha-I is the possibility of working on videos captured by a moving camera. During the acquisition process, the camera speed is adjusted to the subject pace.

## 1.1 Related Work

As new needs arise in the world of video surveillance and security, new datasets are created to test recognition methods. In this section we will show an overview of dataset comparable with the one proposed.

COMPACT [2] is a biometric dataset focused on less-cooperative face recognition. Images are in high resolution but acquired in a fully automated manner. This allow to have real-world degradation like expressions, occlusion, blur etc.

Differently from COMPACT, UBEAR [3] is focused on ear images. Subjects in UBEAR are in movements and under varying lighting conditions. The subject can move the head freely and acquired images are in gray scale. In the outdoor environment we can found datasets as QUIS-CAMPI [4]. In QUIS-CAMPI, subjects are on the move and at about 50 m of distance. For the enroll of data, the same subjects were acquired also in indoor scenario and a 3D model of each subject is also available. There are full body images and, from them, a PTZ camera extracted face images.

In previous example images were captured by a camera, but in the surveillance purpose there are also dataset obtained by drones. In DRONEFACE [5], the authors focused on face tracking. This task became difficult to approach due to the distance between the drone and the subjects. For this reason, they built DRONEFACE, composed of facial images taken from various combinations of distances and heights for evaluating how a face recognition technique works in recognizing designated faces from the air.

Another dataset focused on faces on distance is SALSA [7]. Differently from the previous, in salsa we have a fixed camera network that record subjects in two different modalities, both in indoor environment. The first modality simulate a poster presentation, in which there is a presenter and an audience. The second modality simulate a cocktail party in which subjects are freely to move and interact.

All previous datasets are focused on one, or at most two biometric traits. However, recently, also multibiometric dataset were created.

As an example, MUBIDIUS-I [6] is a multibiometric and multienvironment dataset, acquired by drones and cameras. There are many biometric traits in this dataset as ear, face, iris and full body. Most of the modality are at less distance and with fully cooperative subjects, but there are still videos with less-cooperative subjects at distance in a outdoor environment.

Completely different from the previous ones are then the dataset captured with pose estimation purpose. In order to obtain an accurate ground truth, the most used dataset in this field are like BIWI [8]. BIWI is a face dataset captured by a kinect, that allows faces and 3D models of the subject captured in an indoor environment and in a cooperative mode. There are only 20 subjects in the dataset and only with face informations.

The proposed dataset Gotcha-I brings together all the features of the previous ones. This dataset provide a 3D model of each subjects, despite not having any image captured by depth camera. This make us able to perform pose estimation on real faces with a very precise ground truth as BIWI. As MUBIDIUS-I, our dataset can be used for multiple biometric traits due to the different distances and modalities. As in SALSA, our subjects can move freely in different modalities and as DRONEFACE we are able to perform the tracking of a subjects through different videos. We have also outdoor environments as in QUIS-CAMPI and a 180° videos of the subjects that allows us to perform ear recognition as in UBEAR. Finally our modality is more than less-cooperative than COMPACT because our subjects deliberately try to avoid being filmed.

In Table 1, a list of overall specification of each presented dataset is available, compared with Gotcha-I.

In Table 2 a list of annotation furnished from each compared dataset is present. Very few dataset has 3D models of the subjects and Landmark annotation, by which we mean coordinates of keypoint on the faces or bodies of the subjects.

**Table 1.** Reference datasets with overall specifications. C./N.C. means Cooperative/Non Cooperative Mode

Dataset	Subjects	Biometrics	Environment	Device	C./N.C.
COMPACT	108	Face	Indoor	Camera	No
UBEAR	126	Ear	Indoor	Camera	No
QUIS-CAMPI	320	Full body	Outdoor	Camera	Yes/No
DRONEFACE	11	Face	Outdoor	Camera	Yes/Yes
MUBIDIUS-I	80	Multi	Multi	Multi	Yes/No
SALSA	18	Full body	Multi	Multi	No/Yes
BIWI HEADPOSE	20	Face	Indoor	Camera	Yes/No
Gotcha-I	62	Full body/face	Multi	Camera	Yes/Yes

## 1.2 Security Purpose Applications

Videos captured by mobile camera with different environments are representative of surveillance data. In the last ten years, more and more countries provided their police officers with body worn cameras [9]. Differently from fixed cameras, mobile cameras has the ability to move around big spaces and cover big areas. However,

**Table 2.** Reference datasets with type of annotations

Dataset	3D Models	Landmark annotation
COMPACT	No	No
UBEAR	No	No
QUIS-CAMPI	Yes	No
DRONEFACE	No	No
MUBIDIUS-I	No	No
SALSA	No	Yes
BIWI HEADPOSE	Yes	No
Gotcha-I	Yes	Yes

data as images and videos from mobile cameras are quite different from fixed cameras due to the different point of view during recording [10]. Our dataset is proposed as a starting point to test algorithms operating on that topic and we will introduce some algorithms that could benefit from it.

**Identity Recognition.** A hot topic in surveillance is identity recognition. One of the most used biometric traits for this purpose is the face. In this sense, there are a lot of algorithms in last years that use faces to recognize a subject, both using neural networks [11, 12] or mixed methods [13]. In order to do that, on this type of data may be useful use previously a head pose estimation method to select the most frontal frame. This aim can be reached with various algorithm in literature that works in real time, both with Neural Networks and without [14–16]. The proposed dataset also allows to detect a subject from other biometric traits like ear or iris, as algorithms in [17, 18] performs. This is possible due to the different distance of the subject from the camera during recording. Thanks to the fact that this happens in the same video sequence, we are also able to fuse biometric traits in order to perform multibiometric recognition using different frames for different biometric trait [19–21].

**Traits Classification.** Not only the identity of the subject can be useful in security purpose, often we are interested in some physical traits like the gender, the age or facial characteristics in order to classify many subjects in few time. Regarding the gender recognition, it is often performed using face [22, 23]. However, it is also possible to use videos in which subjects can move freely, also far from camera, using the gait as biometric traits [24, 25] or data collected by mobile devices [26]. It is also interesting how different results are if we consider the cooperativeness or the non cooperativeness of the subjects, labeled in this dataset, as the work in [27]. Using the face or the movement of the subject we can also extract information about their age, as in [29, 30]. This problem is often performed together with the previous one, gender recognition, to extract the most intuitive and general information about the subjects [31]. Once some characteristics are captured we may be able to follow the subjects along different

paths and cameras, as our dataset allow us. This is called tracking [32,33] and it is a very hot topic in security due to the ability to find the same people in different environments, in real time, without the need to know the exact identity of the subject [34]. For this purpose, if we focus our attention on face, there are various characteristics we can use to discriminate subjects, as in [28]. At the same times various classification and clustering algorithms were built in this sense [35–37].

In conclusion, taking a look at identification and classification of subjects in security purpose, our dataset allow users to train, test and perform very different algorithms at the state of the art.

## 2 Gotcha-I

Our proposed dataset stems from the growing need to extract biometrics from video surveillance data and from the need to understand who the user is, where he is and what he is doing. Gotcha-I dataset allows to extract different biometrics: the face, the nose, the mouth, the eyes, the ears and the periocular area. Given the nature of the videos it is also possible to extract behavioral biometrics from gait. Gotcha-I dataset simulates the acquisition of the body worn camera in which a moving subject is acquired by a moving camera. It is available for download at [1].

### 2.1 Content of the Dataset

To simulate real-world conditions, no accessories (clothes, hats or glasses) were controlled, they were left participant dependent. About the procedure followed by each participant, there were two recording procedures: (I) a cooperative mode with the camera where the subject walks and collaborates with the camera watching it during the walk, see Fig. 1 (top-left), and (II) a non cooperative mode where the same subject walks trying to avoid the camera, see Fig. 1 (top-right). The dataset contains a total of 493 videos with an average duration of 4 min, including 62 subjects, 15 women and 47 men in an average age between 18 and 20 years.

In order to be able to create robust systems, several possible scenarios were considered for the previous described procedures. The dataset is composed of 11 different video modes in different environmental and behavioral contexts.

The contents of the dataset are listed below:

- (EC1) indoor with artificial light - cooperative mode;
- (EC2) indoor with artificial light - non cooperative mode;
- (EC3) indoor without any lights but the camera flash - cooperative mode;
- (EC4) indoor without any lights but the camera flash - non cooperative mode;
- (EC5) outdoor with sunlight - cooperative mode;
- (EC6) outdoor with sunlight - non cooperative mode;
- (EC7) 180° head video;
- (EC8) stairs outdoor - cooperative mode;



**Fig. 1.** Some Gotcha-I dataset samples: outdoor sunlight in cooperative mode (top left), indoor with artificial light in non cooperative mode (top right) and indoor with the camera flash in cooperative mode (down).

- (EC9) stairs outdoor - non cooperative mode;
- (EC10) path outdoor - cooperative mode;
- (EC11) path outdoor - non cooperative mode;
- (EC12) derived files attached, detailed in Sect. 2.2.

All the videos have been acquired with the camera of the Samsung S9+ mobile phone; the modes (EC8-EC9-EC10-EC11) have also been acquired with an iPhone 10 and a Samsung Galaxy A5.

**Illumination Differences.** Some real world problems can occur regarding the selected illumination settings. Videos in (EC5-EC6-EC8-EC9-EC10-EC11) were captured outdoors with natural sunlight, Fig. 1 (top-left). Videos in (EC1-EC2-EC7) have been acquired in a room with a white background with the artificial lights on, see Fig. 1 (top-right). Videos in (EC3-EC4) have been acquired in the same room with the lights off and the flash camera on. In these video the use of camera flash can generate blur frames in some sequences Fig. 1 (down); as we can see, in this mode some frames could be blurred increasing the dataset complexity.

**Cooperative and Non-cooperative Mode.** In cooperative video sequences, the subjects look at the camera during the acquisition and follow the camera lens during the motion. In non-cooperative video sequences, the subjects try to avoid the camera during the motion, can be appreciated in Fig. 1 (top-right). This modality is clearly most competitive.

Figure 2-a shows the distance from the neck to the nose in a cooperative video and Fig. 2-b shows this distance in a non-cooperative video of the same subject. We can observe that the cooperative mode exhibits a linear behaviour, while the non cooperative mode behaves irregular. The differences in the regularity of the subjects pose in cooperative and non-cooperative videos have led us to carry out experiments on different methods to further analyze these differences.

**Path and Stairs Outdoor.** In these two modes the videos were acquired from different points of view simultaneously between three different cameras: Samsung S9 +, iPhone 10 and a Samsung Galaxy A5. Furthermore not all subjects are present. These videos have been created specifically to perform re-identification and action recognition algorithms. Our aim is that these sequences simulate a video surveillance camera acquisition, so once the face (or the gait) is acquired it is possible to re-identify it and trace it for the whole journey. Furthermore, the action of “going up the stairs” allows to perform algorithms of action recognition in order to predict if a subject is going up the stairs or walking. Example of different frames extracted from these videos are shown in Fig. 3 .

**180° Head Video.** The facial video sequences were acquired with the most favorable lighting conditions: indoor with lights on. There are 62 sequences, one for each subject, close to less than a meter from the face by rotating the camera around the head 180°: from the left ear to the right ear. The subjects were asked

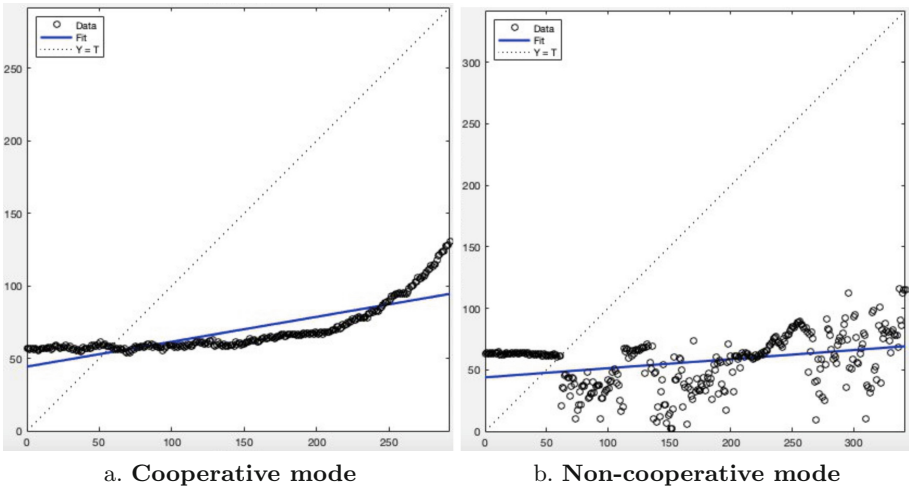


Fig. 2. Head-pose variation sequence for each mode.



to sit on a chair placed in a room with a white panel behind them, the operator then made the video by turning around the subject. This mode has the purpose of acquiring the facial details and consequently, can be used to analyze facial traits that require a high resolution of the image, such as iris, ear, profile, nose, mouth, and periocular area (Fig. 3).



**Fig. 3.** Different outdoor sequences.

## 2.2 Additional Metadata

Additional related information such as 3D-data extracted from videos is included in our dataset. It was possible to extract from each 3D model the pitch yaw and roll rotation of the face for the Head Pose Estimation.

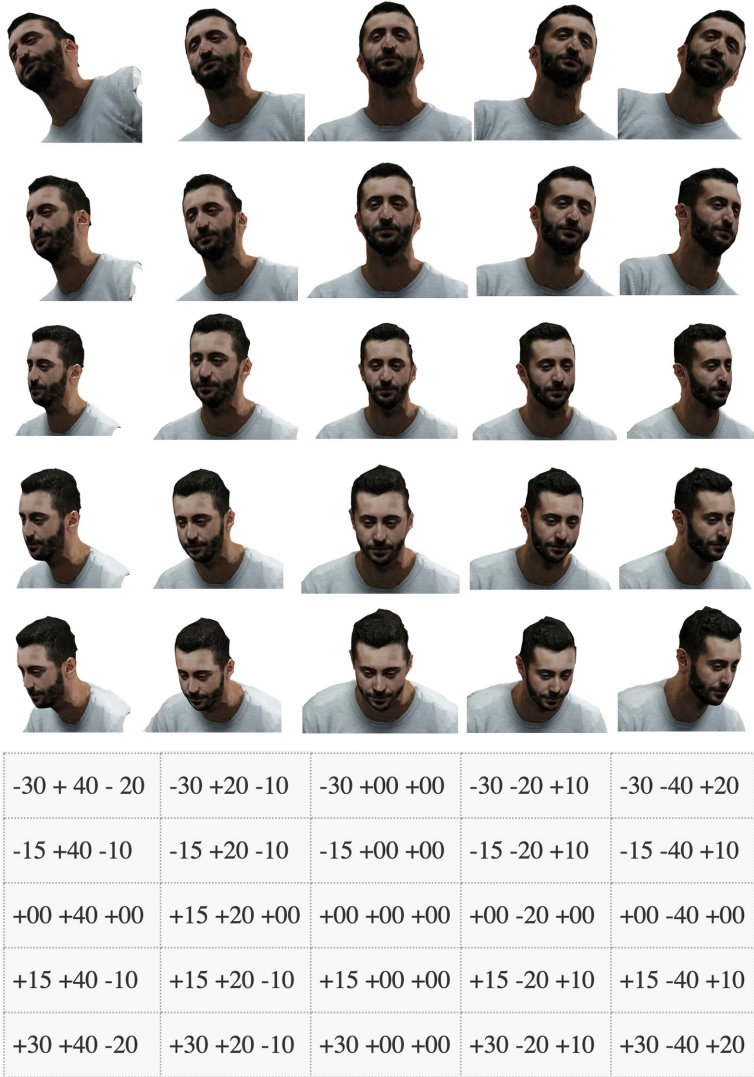
**3D Model and Head Pose Estimation Data.** From the videos in “(EC7) 180° head video” we have reconstructed the 3D model of the head in .obj format available within the derived files attached. From the 3D model of the head adequately elaborated through the Blender software, the images of the head were extracted in all the pitch, yaw and roll poses with 5° deviations in the following ranges of values:

- Pitch ( $-30^\circ$ ;  $+30^\circ$ );
- Yaw ( $-40^\circ$ ;  $+40^\circ$ );
- Roll ( $-20^\circ$ ;  $+20^\circ$ ).

For 62 subjects therefore 137.826 images were extracted. In Fig. 4 there is a subset of 25 images of the head pose estimation of the subject 62.

**Landmark Extraction Data.** For each video frame, except for the videos in “(EC7) 180° head video”, the landmarks of the 2D pose estimation of the body and the 68 landmarks of the face were extracted using the OpenPose software [38]. This data is useful for gait analysis and performance action recognition.





**Fig. 4.** Above some examples of image extracted from 3D model of the subject 62. At the bottom the degrees in pitch, yaw and roll of the head pose estimation corresponding to the position in the table.

### 3 Conclusion

Gotcha-I is a multiview dataset built to meet the needs of surveillance data from Body Worn Cameras. With a total of 62 subjects in 11 different modalities, Gotcha-I results particularly suitable for tasks such as people tracking and recognition. The high definition and full bodies in video allow to perform dif-

ferent type of biometric traits, both physical and behavioral. Compared to the other datasets in literature, Gotcha-I presents a remarkable difference between cooperative and non cooperative modalities, allowing to analyze the response of different state-of-the-art algorithms on this data. Additional contents as the 3D model of each subjects, face and body coordinates and annotated head pose images, make our dataset very versatile in terms of possible testable applications.

## References

1. Gotcha-I dataset. <https://gotchaproject.github.io/>
2. Włodarczyk, M., Kacperski, D., Sankowski, W., Grabowski, K.: COMPACT: biometric dataset of face images acquired in uncontrolled indoor environment. *Comput. Sci.* **20**(1) (2018). <https://doi.org/10.7494/csci.2019.20.1.3020>
3. Raposo, R., Hoyle, E., Peixinho, A., Proença, H.: UBEAR: a dataset of ear images captured on-the-move in uncontrolled conditions. In: 2011 IEEE Workshop on Computational Intelligence in Biometrics and Identity Management (SSCI 2011 CIBIM), Paris, France, 11–15 April, pp. 84–90 (2011)
4. Neves, J., Moreno, J., Proença, H.: QUIS-CAMPI: an annotated multi-biometrics data feed from surveillance scenarios. *IET Biom.* **7**(4), 7 (2018). <https://doi.org/10.1049/iet-bmt.2016.0178>
5. Hsu, H.J., Chen, K.T.: DroneFace: an open dataset for drone research. In: Proceedings of the 8th ACM on Multimedia Systems Conference (MMSys 2017), pp. 187–192. ACM, New York (2017). <https://doi.org/10.1145/3083187.3083214>
6. Di Maio, L., Distasi, R., Nappi, M.: MUBIDUS-I: a multibiometric and multipurpose dataset. In: SITIS 2019 - The 15h International Conference on Signal Image Technology and Internet Based Systems, 26–29 November 2019, Sorrento, Italy (2019)
7. Alameda-Pineda, X., et al.: SALSA: a novel dataset for multimodal group behavior analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* **38**(8), 1707–1720 (2015)
8. Fanelli, G., Dantone, M., Gall, J., Fossati, A., Van Gool, L.: Random forests for real time 3D face analysis. *Int. J. Comput. Vision* **101**, 437–458 (2013)
9. Bromberg, D.E., Charbonneau, É., Smith, A.: Public support for facial recognition via police body-worn cameras: findings from a list experiment. *Gov. Inf. Q.* **37**(1), 101415 (2019)
10. Younis, O., Al-Nuaimy, W., Rowe, F., Alomari, M.H.: Real-time detection of wearable camera motion using optical flow. In: 2018 IEEE Congress on Evolutionary Computation (CEC). <https://doi.org/10.1109/CEC.2018.8477783>
11. Ding, C., Tao, D.: Trunk-branch ensemble convolutional neural networks for video-based face recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **40**(4), 1002–1014 (2018). <https://doi.org/10.1109/TPAMI.2017.2700390>
12. Guo, G., Zhang, N.: A survey on deep learning based face recognition. *Comput. Vis. Image Underst.* **189**, 102805 (2019)
13. Yue, G., Lu, L.: Face recognition based on histogram equalization and convolution neural network. *IEEE Trans. Pattern Anal. Mach. Intell.* **40**(4), 1002–1014 (2018). <https://doi.org/10.1109/TPAMI.2017.2700390>
14. Ranjan, R., Patel, V.M., Chellappa, R.: Hyperface: a deep multi-task learning framework for face detection, landmark localization, pose estimation, and gender recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **41**, 121–135 (2019)

15. Barra, P., Bisogni, C., Nappi, M., Ricciardi, S.: A survey on deep learning based face recognition. *Computer Vis. Image Underst.* **189**, 102805 (2019)
16. Abate, A.F., Barra, P., Bisogni, C., Nappi, M., Ricciardi, S.: Near real-time three axis head pose estimation without training. *IEEE Access* **7**, 64256–64265 (2019)
17. Chowdhury, D.P., Bakshi, S., Sa, P.K., Majhi, B.: Wavelet energy feature based source camera identification for ear biometric images. *Pattern Recogn. Lett.* **130**, 139–147 (2018)
18. Llano, E.G., Vázquez, M.S.G., Vargas, J.M.C., Fuentes, L.M.Z., Acosta, A.A.R.: Optimized robust multi-sensor scheme for simultaneous video and image iris recognition. *Pattern Recogn. Lett.* **101**, 44–51 (2018)
19. Sonal, Singh, A.: Review on multibiometrics: classifications, normalization and fusion levels. In: 2018 International Conference on Advances in Computing and Communication Engineering (ICACCE), 22–23 June 2018. IEEE (2018). <https://doi.org/10.1109/ICACCE.2018.8441727>
20. Bisogni, C., Nappi, M.: Multibiometric score-level fusion through optimization and training. In: 2019 3rd International Conference on Bio-engineering for Smart Technologies (BioSMART). 24–26 April 2019. IEEE (2019). <https://doi.org/10.1109/BIOSMART.2019.8734162>
21. Abate, A.F., Bisogni, C., Castiglione, A., Distasi, R., Petrosino, A.: Optimization of score-level biometric data fusion by constraint construction training. In: Wang, G., El Saddik, A., Lai, X., Martinez Perez, G., Choo, K.-K.R. (eds.) *iSCI 2019. CCIS*, vol. 1122, pp. 167–179. Springer, Singapore (2019). [https://doi.org/10.1007/978-981-15-1301-5\\_14](https://doi.org/10.1007/978-981-15-1301-5_14)
22. Dhomne, A., Kumar, R., Bhan, V.: Gender recognition through face using deep learning. *Proc. Comput. Sci.* **132**, 2–10 (2018)
23. Cerkezi, L., Topal, C.: Gender recognition with uniform local binary patterns. In: 2018 26th Signal Processing and Communications Applications Conference (SIU) (2018). <https://doi.org/10.1109/SIU.2018.8404587>
24. Isaac, E.R., Elias, S., Rajagopalan, S., Easwarakumar, K.S.: Multiview gait-based gender classification through pose-based voting. *Pattern Recogn. Lett.* **126**, 41–50 (2019)
25. Barra, P., Bisogni, C., Nappi, M., Freire Obregon, D., Castrillon-Santana, M.: Gender classification on 2D human skeleton. In: 3rd International Conference on Bio-Engineering for Smart Technologies (BioSMART 2019) (2019). <https://doi.org/10.1109/BIOSMART.2019.8734198>
26. Jain, A., Kanhangad, V.: Gender classification in smartphones using gait information. *Exp. Syst. Appl.* **93**(1), 257–266 (2018)
27. Barra, P., Bisogni, C., Nappi, M., Freire-Obregón, D., Castrillón-Santana, M.: Gait analysis for gender classification in forensics. In: Wang, G., Bhuiyan, M.Z.A., De Capitani di Vimercati, S., Ren, Y. (eds.) *DependSys 2019. CCIS*, vol. 1123, pp. 180–190. Springer, Singapore (2019). [https://doi.org/10.1007/978-981-15-1304-6\\_15](https://doi.org/10.1007/978-981-15-1304-6_15)
28. Liu, Z., Luo, P., Wang, X., Tang, X.: Large-scale celebfaces attributes (CelebA) dataset. The Chinese University of Hong Kong, Multimedia Laboratory (2015)
29. Pakulich, D.V., Yakimov, S.A., Alyamkin, S.A.: Age recognition from facial images using convolutional neural networks. *Optoelectron. Instrument. Data Process.* **55**(3), 255–262 (2019). <https://doi.org/10.3103/S8756699019030075>
30. Iqbal, M.T.B., Shoyaib, M., Ryu, B., Abdullah-Al-Wadud, M., Chae, O.: Directional age-primitive pattern (DAPP) for human age group recognition and age estimation. *IEEE Trans. Inf. Forensics Secur.* **12**(11), 2505–2517 (2017)

31. Rodríguez, P., Cucurull, G., Gonfaus, J.M., Roca, F.X., Gonzalez, J.: Age and gender recognition in the wild with deep attention. *Pattern Recogn.* **72**, 563–571 (2017)
32. Wen, S.-Y., Yen, Y., Chen, A.Y.: Human tracking for facility surveillance. In: Arai, K., Kapoor, S. (eds.) *CVC 2019. AISC*, vol. 944, pp. 329–338. Springer, Cham (2020). [https://doi.org/10.1007/978-3-030-17798-0\\_27](https://doi.org/10.1007/978-3-030-17798-0_27)
33. Dadi, H.S., Pillutla, G.K.M., Makkena, M.L.: Face recognition and human tracking using GMM, HOG and SVM in surveillance videos. *Ann. Data Sci.* **5**(2), 157–179 (2017). <https://doi.org/10.1007/s40745-017-0123-2>
34. Lee, Y.G., Chen, S.C., Hwang, J.N., Hung, Y.P.: An ensemble of invariant features for person reidentification. *IEEE Trans. Circ. Syst. Video Technol.* **27**(3), 470–483 (2017)
35. Anzalone, L., Barra, P., Barra, S., Narducci, F., Nappi, M.: Transfer learning for facial attributes prediction and clustering. In: Wang, G., El Saddik, A., Lai, X., Martinez Perez, G., Choo, K.-K.R. (eds.) *iSCI 2019. CCIS*, vol. 1122, pp. 105–117. Springer, Singapore (2019). [https://doi.org/10.1007/978-981-15-1301-5\\_9](https://doi.org/10.1007/978-981-15-1301-5_9)
36. Wang, P., Su, F., Zhao, Z.: Joint multi-feature fusion and attribute relationships for facial attribute prediction. *2017 IEEE Visual Communications and Image Processing (VCIP)* (2017). <https://doi.org/10.1109/VCIP.2017.8305036>
37. Zhuang, N., Yan, Y., Chen, S., Wang, H., Shen, C.: Multi-label learning based deep transfer neural network for facial attribute classification. *Pattern Recogn.* **80**, 225–240 (2018)
38. Cao, Z., Simon, T., Wei, S.E., Sheikh, Y.: *OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields* (2018)